Expert System for Sentence Recognition

Bipul Pandey, Anupam Shukla, and Ritu Tiwari

Department of Information and Communication Technology,
ABV-Indain Institute of Information Technology and Management, Gwalior, India
{pandeybipul,sushilranjan007,
dranupamshukla,tiwariritu2}@gmail.com

Abstract. The problem of using natural languages as a medium of input to computational system has long intrigued and attracted researchers. This problem becomes especially acute for systems that have to deal with massive amount of data as inputs in the form of sentences/commands/phrase as a large number of such phrases may look vastly different in lexical and grammatical structure but yet convey similar meanings. In this paper, we describe a novel approach involving Artificial Neural Network to sufficiently solve the aforesaid problem for inputs in English language. The proposed system uses Self Organizing Map (SOM) to recognize and classify the input sentences into classes representing phrases/sentences having similar meaning. After Detailed analysis and evaluation, we have been able to reach a maximum efficiency of approximately 92.5% for the system. The proposed expert system could be extended to be used in the development of efficient and robust systems like intelligent medical systems, Systems for Intelligent Web-Browsing, telemarketing and several others which will be able to take text input in the form commands/sentences in natural languages to give suitable output.

Keywords: Knowledge Discovery, Natural Language Reasoning, Machine Learning, Artificial Neural Networks, Self Organizing Maps, Sentence Recognition.

1 Introduction

The incredible advances in the field of data storage have enabled the creation of techniques for making robust, credible and efficient databases and data warehouses. However, such massive data depots make the task of knowledge discovery from them, much more difficult, error-prone and non-scalable. Also, it is desired for many such knowledge discovery systems that they should be able to interact directly with the domain experts/researchers and/or users/input providers who are not usually computationally erudite enough so as to give inputs/commands in the form traditionally required by such systems. They usually guide/inform the system through commands/inputs in one or the other spoken languages. These are either in the form of text or speech. The ability of natural languages to express same sentiment in extensively diverse forms has given rise to expressions which may be lexico-grammatically different but yet may have similar meanings. Hence, the development of efficient and reliable systems which can retrieve knowledge from such large sources with

N. Meghanathan et al. (Eds.): CCSIT 2011, Part III, CCIS 133, pp. 66–75, 2011.

information in the form of natural language phrases/commands, is a difficult task and needs novel and suitably chosen approach.

The proposed system makes use of Self Organizing Map (SOM) for efficiently solving the aforesaid problem of mapping input commands/sentences in natural language (here, english) to similar meaning commands/sentences. A Self Organizing Map is a type of Artificial Neural Network which plots a high dimensional space into low dimensional topographical features. Their ability to save the neighborhood related information in inputs made makes them extremely suitable and successful in knowledge discovery from sources comprising of symbolic strings from natural languages.. The network undergoes unsupervised learning procedure. The proposed System comprises of the data source, Coding module to create patterns of words and sentences, specified accordingly and makes use of Self Organizing Maps for classifying and recognizing words and sentences.

2 The State of Art

With a tremendous increase in data amassing, storage and usage, the field of Knowledge Discovery from data sources has seen a significant and intense interest from researchers. KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [1]. With the traditional knowledge discovery measures of manual analysis and interpretation getting obsolete, new theories and tools for computational discovery of useful information are being searched and analyzed. Knowledge Discovery from data sources derive its strength from diverse fields like pattern recognition, Machine Learning statistics, AI, knowledge acquisition for expert systems and others and hence there are many different approaches for the same [2]. Due to the assortment and complexity of data to be mined and the error prone and usually noise-infested nature of the inputs, designing algorithms for proper and suitable learning of such systems is considered an extremely difficult task [3].

Clustering as a tool for the task of Knowledge mining has been extremely popular. Clustering is defined as the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters) [4]. It has been used widely researched upon and used in varied fields. Clustering has been successfully implemented as a measure for pattern recognition and classification [5]. Information recovery and processing has also been successfully attempted through the use of clustering paradigm [6]. Such varied implementation has led to the development of several algorithms and techniques for data and feature clustering implemented and used sufficiently in dwelling upon the depth that the paradigm provides for computationally heavy and complex tasks involved in knowledge mining [7].

Though there is no dearth of efficient algorithms and techniques being used both in Knowledge Discovery and Clustering, Artificial Neural networks have started to gain a foothold because of their versatility and adaptability to representations presented and are increasingly attracting researchers to use them to find novel yet efficient methods of problems being challenged. Though the problem of pattern analysis and data dredging has been widely discussed and debated for Natural Language Reasoning and several existing techniques have been widely used, yet sufficiently successful implementations are few and far in between. [8] talks about Language Modeling to solve the problem of sentence recognition. This is done via using probabilistic grammar along with a Hidden Markov Identifier. In [9], recognition problem is mathematically formulated as an optimization problem with constraints by introducing sentence structures from the syntactic and semantic considerations in speech recognition. In [10], an algorithm had been proposed to describe a framework for classifier combination in grammar-guided sentence recognition. In [11], an algorithm had been proposed for the recognition of isolated off-line words. The algorithm is based on segment string matching and could do with moderately noisy and error prone data set.

In [12], sentence recognition has been achieved which uses a template based pattern recognition and represents words as a series of diaphone-like segments. In [13], word co-occurrence probability has been used for sentence recognition. The incurred results were also compared with the method using the Context Free Grammar. Hybrid Techniques have also been used for the aforesaid problem. In [14], Hidden Markov Model (HMM) and Neural Network (NN) Model have been combined for the solution. Here, Word Recognition had been using a Tree-Structured dictionary while Sentence Recognition is done using a word-predecessor conditioned beam search algorithm to segment into words and word recognition. In [15], hidden Markov models (HMMs) and associative memories have been used for sparse distributed representations. Binary Hamming Neural Network has been applied to recognize sentences and have been found to sufficiently successful in this regard. The system proposed also takes advantage of greater speed of the Binary Networks to provide a very efficient solution to the problem of sentence recognition [16]. David and Rajsekaran have talked about how Hopfield classifiers can be used as a tool in Pattern Classification [17].

3 Methodology

The system makes use of Self Organizing Maps for word and sentence recognition. The Topology of a Self-Organizing Map is as shown in Fig. 1.

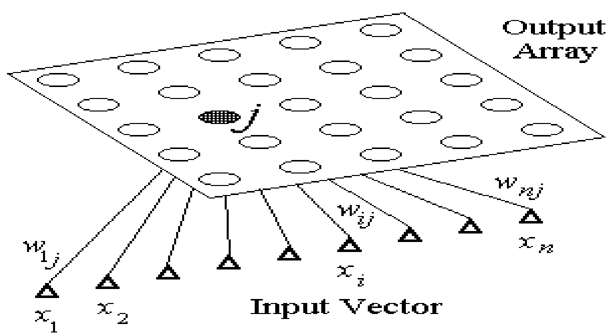

Fig. 1. Topology of a Self Organizing Map

Originally thought as a neural network modelling possible brain structure [18], SOM has been regularly used to analyze/visualize symbolic/text inputs. Their application has led to vastly efficient solution for character recognition as compared to traditionally used approaches [19], [20]. Even from large data sources, their application has vastly improved retrieval of relevant information or knowledge in a very robust manner [21]. Symbolic representations are considered to be highly positional [22]. Hence, the ability of Self Organizing Maps to topologically organize the representations in a cluster are more similar to each other than they are to any of another cluster makes it much more appropriate than over other techniques. The mapping is done so that more frequently occurring data groups are mapped using higher-resolution collections of models.

The Self Organizing Map functions as: First the weights are initialized from N inputs to M output nodes. These weights are very small in value. Now the new input is given to the network. Now the distance (d_j) is calculated from each input pattern to each output node j. This is done according to

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2 \quad . \tag{1}$$

The node j with the minimum distance (d_j) is chosen as the output. Now the weights of the output node j and those in its neighbourhood are updated. The neighbourhood is defined by $NE_j(t)$.

$$w_{ij}(t+1) = w_{ij} + \eta(t)(x_i(t) - w_{ij}(t)) \text{, where} \tag{2}$$

$$j = NE_j(t),\ 0 \le j \le N-1,\ 0 < \eta(t) \,.$$

The gain term $\eta(t)$ decreases with each iteration. This process is kept on repeating till final output is achieved [23].

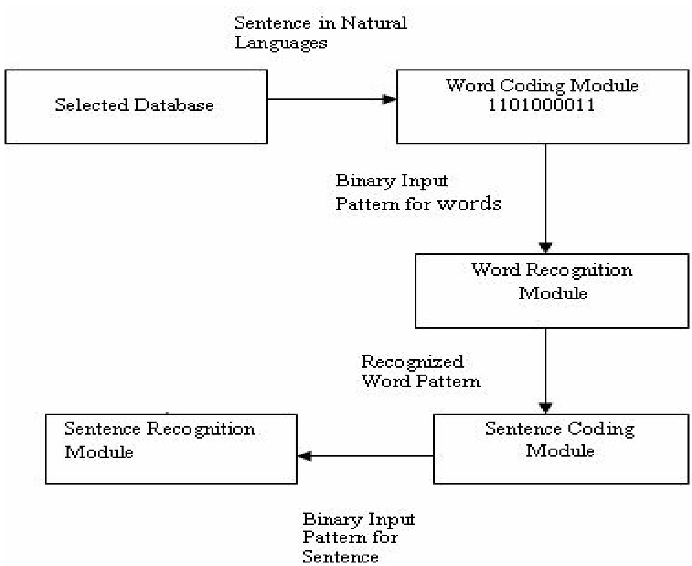

Fig. 2. Architecture of the Proposed System

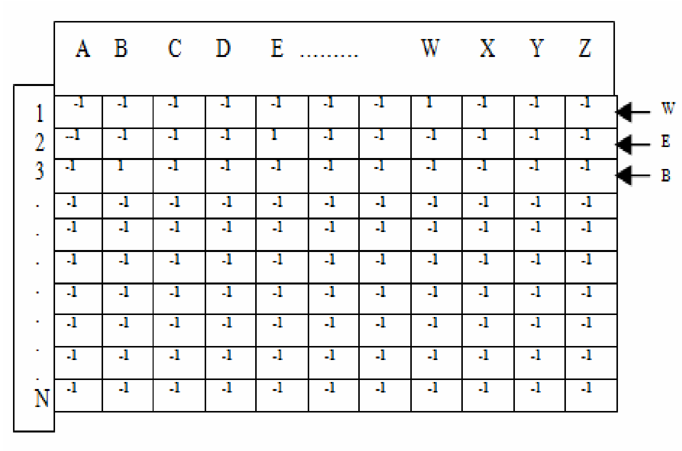

Fig. 3. Format for Binary Coding of Isolated Words

The architecture of the proposed system is as shown in Fig. 2. In the beginning the text data base is used to take inputs. These inputs are sentences/phrases/expressions in natural languages. These inputs are then given to the Word Isolation and Coding Module. Here each sentence/expression is processed by isolating individual words. The words thus extracted are then coded in the form of a matrix. The matrix is of size (N x M), where N = Number of alphabets in the word, &
M = Total English alphabets in word=26.

The words are coded in a binary manner (using 1 & -1). What alphabet is present is represented by the column number in which they are present. For Example, the word is coded as shown in the Fig. 3. The first 'w' is represented by the binary input 1

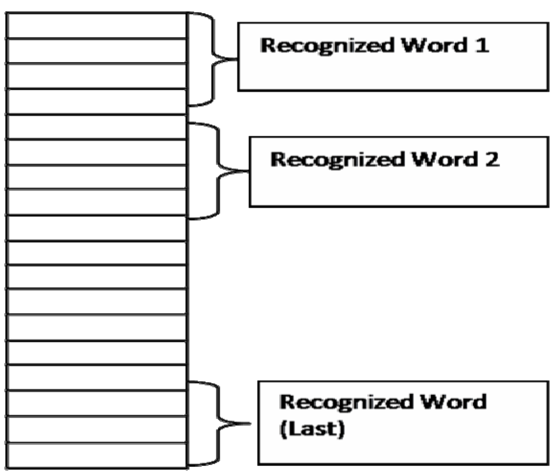

Fig. 4. Format for Binary Coding for Sentences using recognized word pattern in column matrix form

present in (1, 23) position in the matrix. Rest all positions are filled with the binary input -1. Hence, for nth row all positions are -1 except where an alphabet is present.

After this, the coded words are passed on to the word recognition module. This module consists of a Self Organizing Map (SOM) for the recognition of words using clustering paradigm. The input word is mapped on to a specific cluster. The resultant recognized word is the training word representative of the cluster in which the input word has been mapped. The different words recognized for the input sentence pattern are now combined to give a pattern in the sentence coding module which is the input for the sentence recognition module (Fig 4). This is done by appending the output patterns according to the position of the words in the sentence to give the sentence pattern in the column matrix form. These sentence patterns are then sent to the Sentence Recognition Module. This module also comprises of a Self Organizing Map trained with chosen sentence patterns. The network clusters and maps the input pattern to a specific cluster giving result which is the representative training sentence pattern for that cluster. If the pattern falls in a cluster with many or no representative word pattern, then hamming distance is used to find the pattern most similar to it. This output is then matched with the original sentence to check whether the output conveys the same meaning or not. In this way, the input phrases, words or sentences are recognized.

4 Experimental Results

The Database, used for constructing the proposed system, is comprised of 500 sentences, with 50 sentences to be used as training data for the network in the sentence recognition module. The Database here was prepared by taking 500 sentences from a general text regarding common web usage. These sentences were taken as a paragraph in the form of a text file. The different sentences separated by full stops were segregated in the Word Isolation and Coding Module. This module is built using Java Coding. Each sentence had its end marked by a full stop. The code takes advantage of this and separates sentences which start after one full stop and end before the next one. The module then isolates each word per sentence and forms word pattern for each using the matrix format described and outputs each in the form of a text file. The 50 sentences chosen as the training data for the Sentence Recognition Module have their words along with 100 more words taken from the database as the training set for the Word Recognition Module. The remaining words and sentences of the database are chosen as the testing data for the Word Recognition Module and Sentence Recognition Module in the system.

The Sentence pattern of these 50 sentences is formed by appending the word patterns of each word of a sentence as per their position in the sentence to give the sentence pattern of that particular sentence. These patterns are then used as the training set for the Sentence Recognition Module. The remaining sentences of the database are chosen as the testing data for the system. The words of each of the testing sentence are fed to the Word Recognition Module which gives the training word pattern most similar to that word as its output according to the Neural Network used in the module to classify the input patterns. If Self-Organizing Map is used to clssify the input word patterns in the Word Recognition Module, it topologically maps the pattern to a specific cluster. The training word representative (the training word pattern which was mapped to the same cluster) of the cluster is given as the output in this case. If the pattern falls in a cluster with many or no representative word pattern, then hamming distance is used to find the pattern most similar to it.

Table 1. Details of SOM model used

Model	No. of Layers (word recognition module, sentence recognition module)	No. of Neurons (word recognition module, sentence recognition module)	Transfer Function
Self Organizing Maps (SOM)	(single, single)	(393,50)	-

Table 2. Details of BPA model used

Model	No. of Layers (word recognition module, sentence recognition module)	No. of Neurons (word recognition module, sentence recognition module)	Transfer Function
Back Propagation Algorithm (BPA)	(single, single)	(393,500)	transig, purelin

The output patterns of each word are then appended as per its position to give the sentence pattern in the column matrix form as described earlier. The patterns of each of the sentence are then given to the Sentence Recognition Module where the training sentence pattern most similar to the input pattern is given as its output according to the Neural Network used as a classifier in the module. If Self-Organizing Map is used here, it will map the pattern based on topological similarity to a cluster. The training sentence representative (the training sentence pattern which was mapped to the same cluster) of the cluster is given as the output in this case. If the pattern falls in a cluster with many or no representative word pattern, then hamming distance is used to find the pattern most similar to it. The output of each testing sentence is then checked with the original testing sentence to find out where the output is similar in meaning it or not.

Table 3. Details of RBF model used

Model	No. of Layers (word recognition module, sentence recognition module)	No. of Neurons (word recognition module, sentence recognition module)	Transfer Function
Radial Basis Function (RBF)	(single, single)	(393,500)	Radial basis

Along with Self-Organizing Maps, we also implemented Back Propagation Algorithm (BPA) and Radial Basis Function (RBF) for word and sentence recognition and compared the individual results achieved. The details of each Artificial Neural Network used are given in Table 1, Table 2 and Table 3.

The proposed system using Self Organizing Map gave an accuracy of 95.5%approximately. On the other hand, the efficiency achieved using Back Propagation Algorithm (BPA) and Radial Basis Function (RBF) was found to be just 68% and 74% respectively. Hence, the results show that the proposed system using

Table 4. Experimental Results Achieved using SOM

Model Used	Total No. of Sentence Pattern In Database	No. of Training Sentence Patterns	No. of Training Word Pattern	Performance
Self Organizing Maps	500	50	393	430/450

Table 5. Experimental Results Achieved using BPA

Model Used	Total No. of Sentence Pattern In Database	No. of Training Sentence Patterns	No. of Training Word Pattern	Performance
Back Propagation Algorithm	500	50	393	306/450

Table 6. Experimental Results Achieved using RBF

Model Used	Total No. of Sentence Pattern In Database	No. of Training Sentence Patterns	No. of Training Word Pattern	Performance
Radial Basis Function	500	50	393	333/450

Self-Organizing Maps is able to solve this problem of natural language reasoning in a reasonably efficient manner in a more efficient manner as compared to other conventional methods.

5 Conclusion and Future Works

The approach to the problem of Natural Language Reasoning proposed in this paper proves is a novel one. The experimental results also prove the contention that the system is able to give an extremely high level of performance and accuracy in deciphering patterns in natural languages. The approach proposed in this paper is extremely scalable and flexible for text input sources and could also be extended for speech and image input sources too. The proposed system could be extended and implemented for development of specific expert systems like for intelligent medical computational system, intelligent detection system for emotions in expressions and likewise.

The overall approach is a novel attempt towards making a system that could be intelligent enough to take over certain human task of language oriented text processing and could be robust enough in order to work to provide efficient solutions to such tasks.

References

1. Fayyad, U.M., Shapiro, G.P., Smyth, P.: From Data Mining to Knowledge Discovery: An Overview. In: Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park (1996)
2. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Addison Wesley (2005)
3. Dunham, M.H.: Data Mining Introductory and Advanced Topics. Prentice-Hall, Englewood Cliffs (2003)
4. Jain, A.K., Murthy, M.N., Flynn, P.L.: Data Clustering: A Review. ACM Computing Surveys 31(3), 264–323 (1999)
5. Anderberg, M.R.: Cluster Analysis for Application. Academic Press, London
6. Rasmussen, E.: Clustering algorithms in Information Retrieval: Data Structures and Algorithms. W. B. Frakes, R. Baeza-Yates, Prentice-Hall Publishers (1992)

7. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
8. Jones, G.J.F., Wright, J.H., Wrigley, E.N., Carey, M.J.: Isolated-Word Sentence Recognition Using Probabilistic Context Free Grammar. In: Systems and Applications of Man Machine Interaction Using Speech I/O, IEEE Colloquium, pp. 13/1–13/5 (1999)
9. Shirai, K.: Feature extraction and sentence recognition algorithm in speech input system. In: Proceedings of 4th International Joint Conference on Artificial Intelligence, vol. 1. Morgan Kaufmann, San Francisco
10. Jiang, X., Yu, K., Bunke, H.: Classifier combination for grammar-guided sentence recognition. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 383–392. Springer, Heidelberg (2000)
11. Favata, J.T.: General Word Recognition: Using Approximate Segment-String Matching. In: Proceedings of 4th International Conference on Document Analysis and Recognition, pp. 92–96 (1997)
12. Rosenberg, A.E.: Connected Sentence Recognition Using Diphone-like Templates. In: International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 473–476 (1998)
13. Murase, I., Nakagawa, S.: Sentence Recognition Method Using Word Co-occurrence Probability and Its Evaluation. In: Proceedings of ICSLP, pp. 1217–1220 (1990)
14. Marukatat, S., Artikres, T., Gallinari, P., Dorizzi, B.: Sentence Recognition through Hybrid Neuro-Markovian Modeling. In: Proceedings of Sixth International Conference on Document Analysis and Recognition, ICDAR 2001 (2001)
15. Markert, H., Kayikci, Z.K., Palm, G.: Sentence understanding and learning of new words with large-scale neural networks. In: Prevost, L., Marinai, S., Schwenker, F. (eds.) ANNPR 2008. LNCS (LNAI), vol. 5064, pp. 217–227. Springer, Heidelberg (2008)
16. Majewski, M., Zurada, J.M.: Sentence Recognition Using Artificial Neural Network. Knowledge-Based Systems 27(1) (2008)
17. David, V.K., Rajasekaran, S.: Pattern Recognition Using Neural and Functional Networks. Springer, Heidelberg (2008)
18. Ritter, H., Kohonen, T.: Self –Organizing Semantic Maps. In: Biological Cybernetics, pp. 241–254. Springer, Heidelberg (1989)
19. Liou, C., Yang, H.: HandPrinted Character Recognition Based on Spatial Topology Distance Measurement. IEEE Transaction on Pattern Analysis and Machine Intelligence 18(9) (1996)
20. Venkatesh, J., SureshKumar, C.: Tamil Handwritten Character Recognition Kohonen's Self-Organizing Maps. International Journal of Computer Science and Network Secuirty 9(12) (2009)
21. Saarikoski, J., Lourikkala, J.: A study of the use of Self-Organizing Maps in Information Retrieval. Journal of Documentation 65(2), 304–322 (2009)
22. Lippman, R.: An Introduction to Computing with Neural Nets. IEEE Transactions on Acoustic, Speech, and Signal Processing, IEEE Signal Processing Society, Piscataway 4(3), 4–22 (1987)
23. Erwin, E., Obermayer, K., Schulten, K.: Self-organizing Maps: ordering, convergence properties and energy functions. In: Biological Cybernetics. Springer, Heidelberg (1992)